THE ROYAL SOCIETY

doi 10.1098/rstb.2000.0758

The structural basis of protein folding and its links with human disease

Christopher M. Dobson

Oxford Centre for Molecular Sciences, New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT, UK (chris.dobson@chem.ox.ac.uk)

The ability of proteins to fold to their functional states following synthesis in the intracellular environment is one of the most remarkable features of biology. Substantial progress has recently been made towards understanding the fundamental nature of the mechanism of the folding process. This understanding has been achieved through the development and concerted application of a variety of novel experimental and theoretical approaches to this complex problem. The emerging view of folding is that it is a stochastic process, but one biased by the fact that native-like interactions between residues are on average more stable than non-native ones. The sequences of natural proteins have emerged through evolutionary processes such that their unique native states can be found very efficiently even in the complex environment inside a living cell. But under some conditions proteins fail to fold correctly, or to remain correctly folded, in living systems, and this failure can result in a wide range of diseases. One group of diseases, known as amyloidoses, which includes Alzheimer's and the transmissible spongiform encephalopathies, involves deposition of aggregated proteins in a variety of tissues. These diseases are particularly intriguing because evidence is accumulating that the formation of the highly organized amyloid aggregates is a generic property of polypeptides, and not simply a feature of the few proteins associated with recognized pathological conditions. That such aggregates are not normally found in properly functional biological systems is again a testament to evolution, in this case of a variety of mechanisms inhibiting their formation. Understanding the nature of such protective mechanisms is a crucial step in the development of strategies to prevent and treat these debilitating diseases.

Keywords: folding mechanisms; energy landscapes; aggregation; folding diseases; amyloid fibrils; ageing

1. PROTEIN FOLDING AND MISFOLDING

A living organism may contain as many as 100 000 different types of protein. Following synthesis on the ribosome, each protein molecule must fold into the specific conformational state that is encoded in its sequence in order to be able to carry out its biological function. How this process happens is one of the most fascinating and challenging problems in structural biology (Dobson & Fersht 1995). In the cell, folding takes place in a complex and highly crowded environment, and the folding process is aided by a range of auxiliary proteins (Gething & Sambrook 1992; Ellis & Hartl 1999). These proteins include molecular chaperones, whose main role is to protect the incompletely folded polypeptide chain from non-productive interactions, particularly those that result in aggregation, and folding catalysts, whose job is to speed up potentially slow steps in the folding process such as those associated with the isomerization of peptidyl–prolyl bonds and the formation of disulphide bonds. However, it is evident that the code for folding is contained within the amino-acid sequence of the protein itself because it has been shown that proteins can reach their correct folded structure *in vitro* in the absence of any auxiliary factors, providing that appropriate conditions can be found (Anfinsen 1973). The questions of how the fold is encoded in the sequence, and how the process of folding takes place, are at last beginning to be answered in a credible manner. Progress in this area has come about as a consequence of novel experimental strategies to probe the structural transitions that take place during folding *in vitro*, and of innovative theoretical studies designed to simulate these transitions (Dobson & Karplus 1999). Perhaps of greatest importance has been the fact that these approaches have been brought together in a synergistic manner to advance our fundamental understanding of this highly complex process.

In vivo, the beginning of the folding process is the nascent chain as it emerges from the ribosome. *In vitro*, folding begins from a fully formed polypeptide chain that has been unfolded, usually by addition of a chemical denaturant such as urea. In both cases the polypeptide chain is highly disordered before folding is initiated. In the extreme case, approximated rather well for many proteins in denaturants, the protein is said to be in a 'random coil' state, where only local steric interactions constrain the conformations adopted by the molecule (Shortle 1996; Smith *et al.* 1996). In order to achieve the native state encoded by the fold, the protein molecule has to find its way to this unique conformation rather than one of the countless alternatives. Contemplation of this problem gave rise to the 'Levinthal paradox', which can

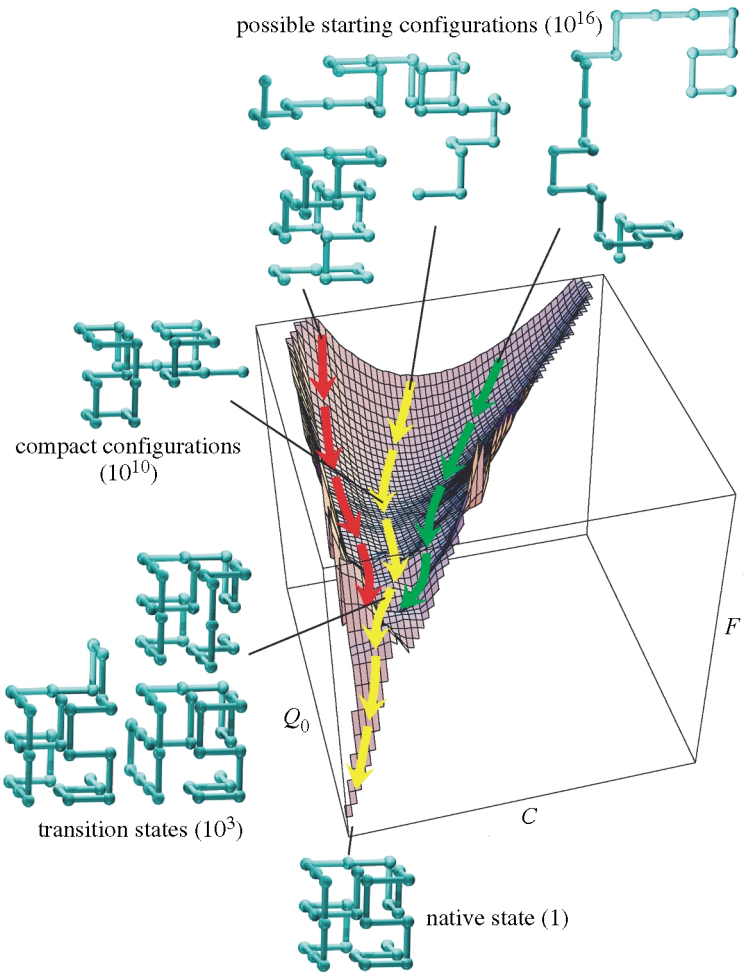

Figure 1. Free-energy (F) surface of a 27-mer model protein as a function of the number of native contacts (Q_0) and the total number of (native and non-native) contacts (C) obtained by sampling the accessible configuration space using Monte Carlo simulations. The yellow trajectory shows the average path traced by structures in 1000 independent trials that each began in a different random conformation. The other two trajectories (green and red) show a range of two standard deviations around the average and are thus expected to include *ca*. 95% of the trajectories. The structures illustrate the various stages of the reaction. From one of the 10^{16} possible random starting conformations, a folding chain collapses rapidly to a disordered globule. It then makes a slow, non-directed search among the 10^{10} semi-compact conformations for one of the approximately 10^3 transition states that lead rapidly to the unique native state. (From Dinner *et al*. 2000.)

be stated most simply as the fact that it would take an almost infinite time for even a small polypeptide chain to search all possible conformations to find the correct (lowest energy) structure, yet real proteins fold rapidly, frequently in less than 1 s (Karplus 1997).

The solution to this apparent paradox has emerged recently through the consideration of the energy surfaces or 'landscapes' on which the folding reaction occurs (Bryngelson *et al*. 1995; Dill & Chan 1997; Dobson *et al*. 1998; Dinner *et al*. 2000). This approach recognizes that folding should not be considered as being analogous to a small molecule reaction, such as breaking and forming a small number of single, strong covalent bonds, but as a biased search on an energy surface that is generally rather flat, i.e. where there are many conformations of similar energy. Such a surface arises because the conformation of a protein is determined by a very large number of relatively weak non-covalent interactions, such as hydrogen bonds and hydrophobic interactions. The bias in folding arises because on average the 'correct', i.e. native-like, contacts between residues are more stable than 'incorrect', or non-native, interactions in any protein that can fold successfully (Dinner *et al*. 2000). On this statistical or 'new view' of protein folding, a given polypeptide chain within the ensemble of conformations making up the denatured state samples only a small number of conformations during its biased search of conformational space, leading it to increasingly lower energy (figure 1). This process is, however, a stochastic one, and different members of the initial conformational ensemble form their stabilizing contacts in very different orders. The energy surface is itself determined from the sequence, and the key feature of any sequence that can fold successfully is that the shape of its energy surface is such that all, or at least the large majority, of the folding routes, or trajectories, lead efficiently to the single lowest energy conformation, the native state.

Considerable progress on the calculation of the surfaces appropriate to protein folding has come about as a result of theoretical simulations of the folding process. Because of the complexity of the problem and the limitations of present computer power, these studies have largely used simplified models of proteins. Such models need to be sophisticated enough to encompass key features of protein folding, e.g. there is a Levinthal paradox, but simple enough to allow many simulations to be carried out in a reasonable time (Dobson *et al*. 1998). Although molecular dynamics simulations, in which all the atoms in the protein are explicitly defined, are becoming viable for this purpose, the most common type of model is a 'lattice model'. A lattice model represents a protein as a string of beads that interact with one other according to a set of simplified potentials (figure 1). The folding process can then be simulated by using Monte Carlo methods in which moves of the chain are biased towards those that result in lower energies. Such methods have been extremely important in establishing the fundamental principles of folding, i.e. how a protein could fold. In order to

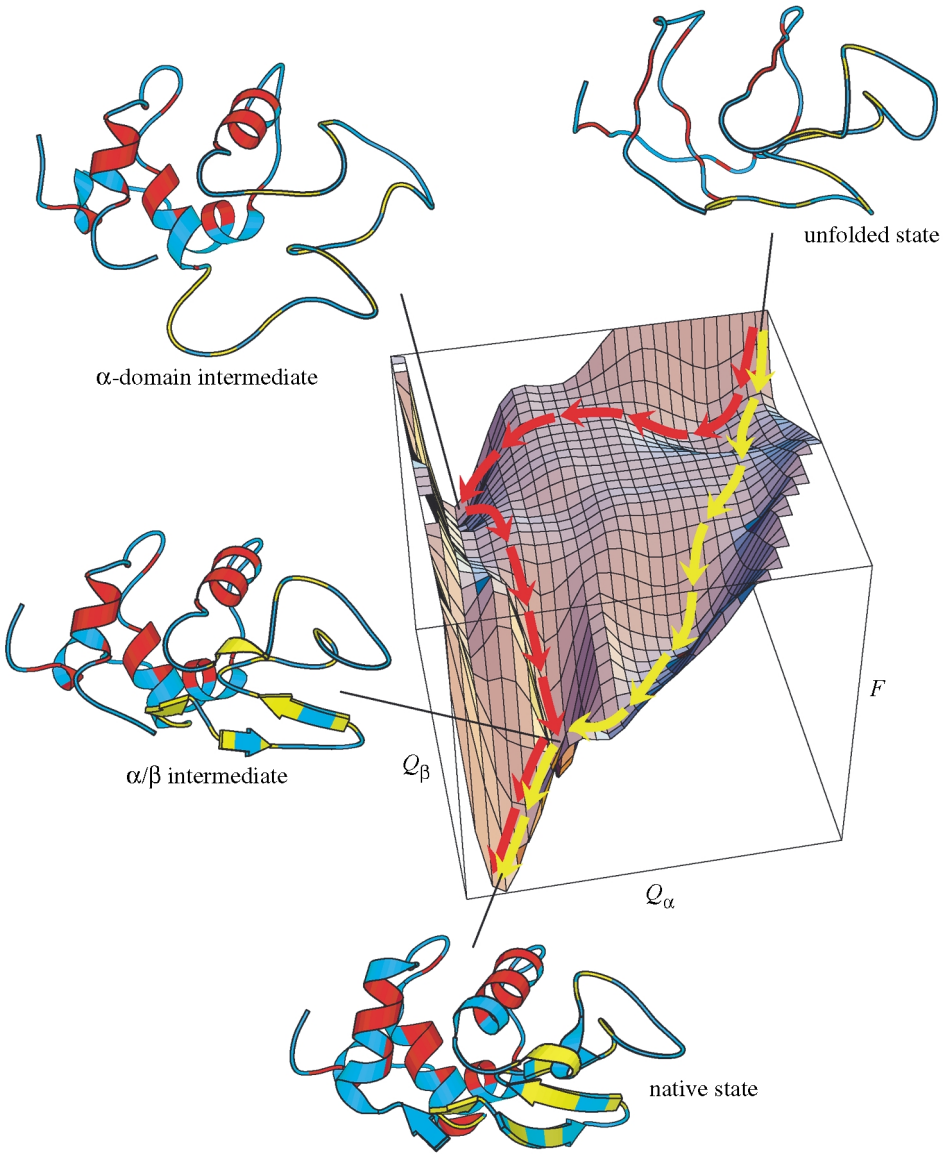

Figure 2. Schematic free-energy (F) surface representing features of the folding of hen lysozyme (a protein of 129 residues whose structure consists of two domains denoted α and β). Q_α and Q_β are the numbers of native contacts in the α and β domains. The yellow trajectory represents a 'fast track' in which the α and β domains form concurrently and populate the intermediate (labelled α/β) only transiently. The red trajectory represents a 'slow track' in which the chain becomes trapped in a long-lived intermediate with persistent structure in only the α domain; further folding from this intermediate involves either a transition over a higher barrier, or partial unfolding to enable the remainder of the folding process to occur along the fast track. Residues whose amide hydrogens are protected from solvent exchange in the native structure (as assessed by NMR) are coloured red (α domain) or yellow (β domain); all others are blue. In each case, regions indicated to be native-like by monitoring the development of hydrogen exchange protection during kinetic refolding experiments are drawn as ribbon representations of the native secondary structure elements (α-helices and β-sheets). (From Dinner *et al.* 2000.)

discover how an actual sequence does fold, however, experimental data need to be obtained to which the simulations can be related.

Investigating the folding of a protein by experimental methods is challenging, both because of the magnitude and rate of the conformational changes that occur, and because of the extreme heterogeneity of the ensembles of conformations that exist at all but the very last stages in the folding process. There are two main approaches that have been developed to overcome these problems. The first is to use biophysical techniques capable of monitoring the properties of a molecular ensemble as folding progresses (Evans & Radford 1994; Plaxco & Dobson 1996). Because of the rapidity of the folding process, such methods usually need to be applied in stopped or quenched flow mode and a number of methods need to be used in concert to map different conformational properties of the ensemble. Thus, for example, far ultraviolet circular dichroism (UV CD) can be used to monitor the evolution of secondary structure during a folding reaction, whilst near UV CD or fluorescence can be used to follow the development of tertiary interactions. An increasing number of techniques is being developed for this purpose on increasingly short time-scales (Callender *et al.* 1998). Of particular importance, because of its ability to define structure at the level of individual residues, is nuclear magnetic resonance (NMR) spectroscopy and significant advances have recently been reported in the application of this technique to study folding (Dobson & Hore 1998; Dyson & Wright 1998).

The second experimental approach to studying the mechanism of folding is to use the methods of protein engineering (Matouschek *et al.* 1989; Fersht 1999). This is a particularly important strategy because it is able to investigate the nature of the transition state for a folding reaction. The essence of the method is to examine the effects on the folding and unfolding kinetics consequent upon mutations of individual residues in the sequence. This approach has been applied in detail to the study of a series of small proteins that fold with apparent two-state kinetics, and provides dramatic evidence that the rate-determining steps in folding involve the formation of native-like contacts around a small number of key residues. This finding has provided compelling evidence for a 'nucleation-condensation' mechanism of folding whereby the majority of the protein structure forms rapidly once a

nucleus of key interactions has formed (Fersht 1997). As well as providing information about the folding of specific proteins, comparison of the transition states of different proteins is beginning to provide insight into the determinants of the folding process. In particular, the transition states of proteins appear to be similar in proteins of similar native state topology, even if the specific sequences are unrelated, and the rates of folding can be correlated with the pattern of the inter-residue contacts formed within different structures (Plaxco *et al.* 1998; Chiti *et al.* 1999*a*; Baker 2000). Conclusions of this type strongly support the concept of a common fundamental mechanism for folding, and are highly encouraging that some relatively straightforward principles will emerge to link a sequence to the structure that it encodes.

The folding of small proteins, typically those of less than about 100 residues, appears to be limited by the time required to search for the crucial interactions that are needed to permit rapid progression to the native structure. For larger proteins, however, the folding process is typically more complex, and is usually associated with the population of one or more partially folded intermediate states. The reason for this is likely to be that larger polypeptide chains have a greater tendency to collapse to a compact state in aqueous solution, even if the contacts between residues in such a state are not those associated with the completely native structure. In such a situation, the rate-limiting steps in folding can be the reorganization of inter-residue interactions within a more-or-less disordered collapsed state (Dobson *et al.* 1998). Within such a state the barriers to reorganization are likely to be much greater than in extended structures, and indeed may involve the breaking of non-native contacts prior to the formation of more stable native ones. This situation can give rise to complex kinetics, and also to distinct heterogeneity in the rate at which different molecules are able to achieve the native structure. In other words, some molecules may form in the initial stages of folding sets of contacts between residues that allow the remainder of the structure to form rapidly. Others, in the statistical process, may form contacts that subsequently generate intermediates that must be substantially reorganized before the complete set of native-like interactions can be formed. Such behaviour is seen both in simulations and in experiments. Indeed, in the case of lysozyme (figure 2), a protein we have studied in great detail, not only is heterogeneity of folding observed under a variety of circumstances, but there is also evidence that this may arise in part from the nucleation of the folding reaction occurring independently in different regions of the structure (Dobson *et al.* 1994; Thirumalai & Klimov 1999; Morozova-Roche *et al.* 1999; Dinner *et al.* 2000).

As the size and complexity of proteins increase, therefore, the folding process becomes more complex. Intermediates with only partially formed structures can be populated and have significant lifetimes. In addition, events that may be termed 'misfolding' may take place during the search for the stable native-like contacts between residues. That such complexities are seen even in the benign environment of a dilute solution of a pure protein suggests that they are even more likely to occur in the crowded environment of the cell. Undoubtedly, molecular chaperones are able to mitigate some of the consequences of this complex behaviour and provide some protection for the incompletely folded chain (Ellis & Hartl 1999). But the idea that proteins can misfold, or fold to intermediates that may undergo undesirable reactions such as aggregation, provides insight into potential problems that can arise during folding even in the best designed environments. Folding and unfolding are also now known to be coupled to many of the key events in the functioning of a biological system, including translocation of proteins across membranes, protein trafficking, secretion of extracellular proteins, and the control and regulation of the cell cycle (Radford & Dobson 1999). Thus, the failure of proteins to fold—or to remain folded under physiological conditions—is likely to cause malfunctions and hence disease (Thomas *et al.* 1995; Dobson 1999*a*). Indeed, an increasing number of diseases is now linked to phenomena that can loosely be described as 'misfolding'; a selection of these is given in table 1.

Table 1. *Representative protein folding diseases*

(ER, endoplasmic reticulum. Data from Tan & Pepys (1994); Thomas *et al.* (1995); Kelly (1998); Lansbury (1999).)

disease	protein	site of folding
hypercholesterolaemia	low-density lipoprotein receptor	ER
cystic fibrosis	cystic fibrosis transmembrane regulator	ER
phenylketonuria	phenylalanine hydroxylase	cytosol
Huntington's disease	huntingtin	cytosol
Marfan syndrome	fibrillin	ER
osteogenesis imperfecta	procollagen	ER
sickle cell anaemia	haemoglobin	cytosol
α1-antitrypsin deficiency	α1-antitrypsin	ER
Tay–Sachs disease	β-hexosaminidase	ER
scurvy	collagen	ER
Alzheimer's disease	β-amyloid/presenilin	ER
Parkinson's disease	α-synuclein	cytosol
scrapie/Creutzfeldt–Jakob disease	prion protein	ER
familial amyloidoses	transthyretin/lysozyme	ER
retinitis pigmentosa	rhodopsin	ER
cataracts	crystallins	cytosol
cancer	p53	cytosol

2. PROTEIN AGGREGATION AND AMYLOID DISEASES

Among the diseases in table 1 are those that are associated with the deposition of proteinaceous aggregates in a variety of organs such as the liver, heart and brain (Tan & Pepys 1994; Prusiner 1997; Kelly 1998; Lansbury 1999). Many of these diseases are described as 'amyloidoses' because the aggregated material stains with dyes such as Congo red in a manner similar to starch (amylose), the aggregates are referred to as 'amyloid' and the typical fibrous structures (figure 3) as 'amyloid fibrils'. A list of known amyloid diseases is given in table 2, along with the protein component that is associated with the extracellular aggregates in each case (Sunde *et al.* 1997). It is

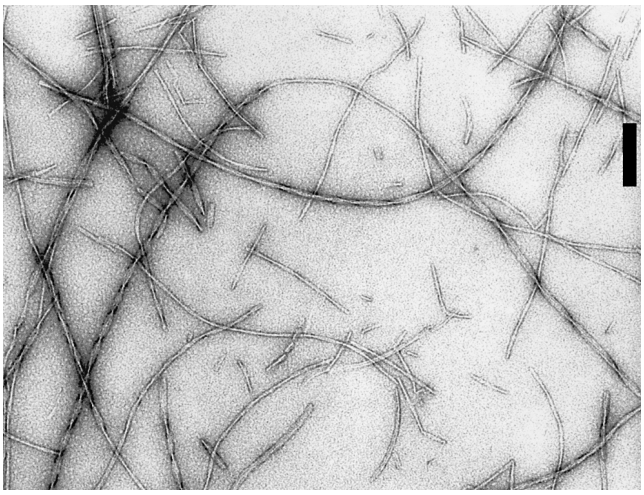

Figure 3. Amyloid fibrils from the Ile56Thr variant of human lysozyme produced by transmission electron microscopy. Scale bar, 200 nm. (From Morozova-Roche *et al*. 2000.)

Table 2. *Fibril protein components and precursors in amyloid diseases*

(Data from Sunde *et al*. (1997).)

clinical syndrome	fibril component
Alzheimer's disease	Aβ peptide, 1-42, 1-43
spongiform encephalopathies	full length prion or fragments
primary systemic amyloidosis	intact light chain or fragments
secondary systemic amyloidosis	76-residue fragment of amyloid A protein
familial amyloidotic poly neuropathy I	transthyretin variants and fragments
senile systemic amyloidosis	wild-type transthyretin and fragments
hereditary cerebral amyloid angiopathy	fragment of cystatin-C
haemodialysis-related amyloidosis	β2-microglobulin
familial amyloidotic polyneuropathy II	fragments of apolipoprotein A-I
Finnish hereditary amyloidosis	71-residue fragment of gelsolin
type II diabetes	fragment of islet-associated polypeptide
medullary carcinoma of the thyroid	fragments of calcitonin
atrial amyloidosis	atrial natriuretic factor
lysozyme amyloidosis	full length lysozyme variants
insulin-related amyloid	full length insulin
fibrinogen α-chain amyloidosis	fibrinogen α-chain variants

evident that these diseases include many of the most debilitating conditions in modern society, particularly those associated with ageing such as type II diabetes and Alzheimer's disease. Some are familial, some associated with medical treatment (e.g. haemodialysis) or infection (the prion diseases), and some are sporadic (e.g. most forms of Alzheimer's). Many of the diseases (such as the amyloidoses associated with the protein transthyretin) can be found in both sporadic and familial forms. In addition to these diseases, there are others—notably Parkinson's and Huntington's diseases (Lansbury 1999; Perutz 1999)—that appear to involve very similar aggregates but which are intracellular not extracellular and are not therefore included in the strict definition of amyloidoses. Many of the deposits include proteins additional to those primarily involved in the fibrillar structures. In addition, there is considerable debate as to whether the fibrillar aggregates themselves give rise to the clinical manifestations of the disease, particularly in the case of the brain pathologies such as Alzheimer's and Creutzfeldt–Jakob diseases, and it has been suggested that precursor aggregates of the fibrils may be the primary cause of the diseases, perhaps instigating the destruction of neurons (Lansbury 1999). In systemic amyloidoses, however, it seems likely that the sheer volume of material involved, sometimes kilogram quantities, by itself can disrupt organs such as the liver and the spleen to cause them to malfunction (Tan & Pepys 1994).

Among the proteins linked with amyloidosis is lysozyme, the protein whose folding we have studied in particular depth. Our studies led to the idea that this protein would be an exciting one to choose to try to understand at the molecular level the nature of the 'misfolding' transition that converts the protein from a soluble to a fibrillar structure (Booth *et al*. 1997). One of the striking characteristics of the amyloid diseases is that the fibrils associated with all of them are very similar in their overall properties and appearance (Sunde & Blake 1997). The fibrils are typically long (often several micrometres), unbranched and *ca*. 10 nm in diameter (figure 3). They have a variety of tinctorial properties, notably staining with Congo red and exhibiting a green birefringence under polarized light. A range of experiments, particularly X-ray fibre diffraction, indicates that the fibrils have extensive β-sheet character, and that these sheets run perpendicular to the fibril axis to generate what is described as a cross-β structure (Sunde & Blake 1997). This observation is remarkable in view of the fact that the soluble native forms of the proteins associated with these diseases vary very considerably in their nature. Some proteins are large, some small, some are largely helical, some largely sheet. Some are intact in the fibrous form, others are at least partially degraded. Some are cross-linked with disulphide bonds and some are not. This similarity of the fibrillar forms of the proteins prompted the proposal that there are strong similarities in the inherent structure of the amyloid fibrils and in the mechanism by which they are formed (Sunde & Blake 1997; Dobson 1999*a*). Thus the study in depth of the relationship between the folding and 'misfolding' of one system could have very general value in understanding this whole class of diseases.

One of the very important observations for this proposal was that the fibrillar forms of many of the disease-related proteins could be generated *in vitro*. In the case of fibrils formed from peptides (often fragments of larger proteins) that are largely unstructured in solution, such fibrils typically form under a wide range of solution conditions. In the case of fibrils formed from intact globular proteins, however, the fibrils typically form under conditions under which the native state is somewhat destabilized (Kelly 1998; Chiti *et al*. 1999*b*). Thus in the case of the two known disease-related human lysozyme

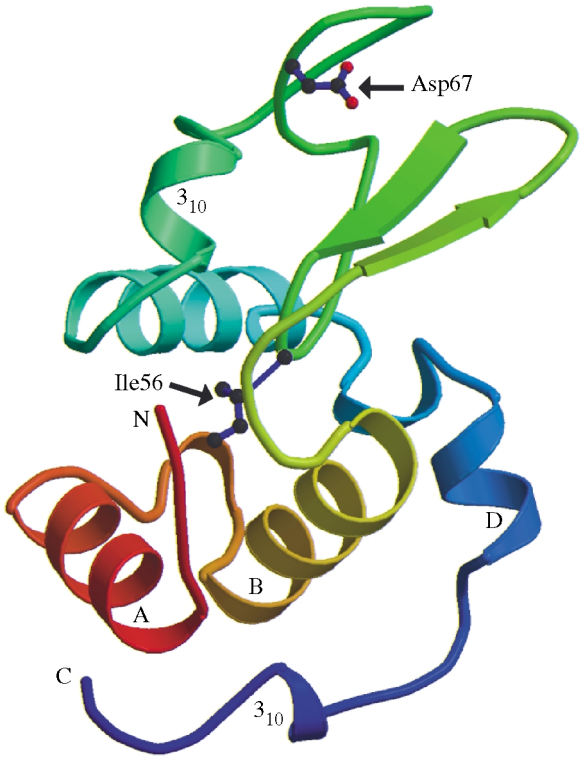

Figure 4. Schematic view of the structure of human lysozyme. The major elements of secondary structure are labelled; the four disulphide bonds are not shown in this representation. The amyloidogenic mutations Ile56Thr and Asp67His are indicated.

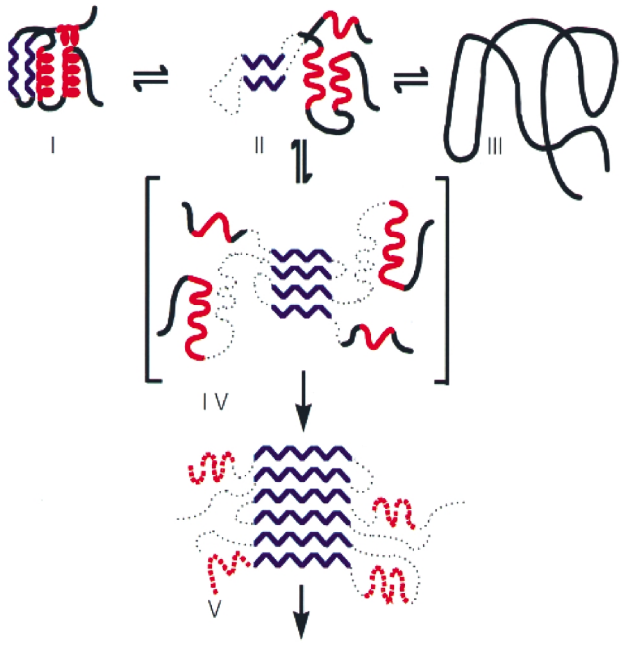

Figure 5. Proposed mechanism for lysozyme amyloid fibril formation. Blue, β-sheet structure: red, helical structure; dotted lines, undefined structure. A partially folded form of the protein (ii) self-associates through the β domain (iv) to initiate fibril formation. This intermediate provides the template for further deposition of protein and for the development of the stable, mainly β-sheet, core structure of the fibril (v). The undefined regions in (v) represent the possibility that not all of the polypeptide sequence is involved in the cross-β structure. The nature of this residual structure in (v) is not known, and the figure is not intended to represent any defined secondary structural type. (From Booth *et al*. 1997.)

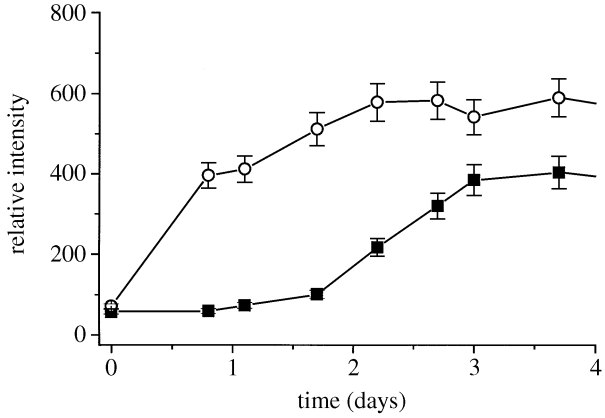

Figure 6. Time dependencies of amyloid fibril formation of the Ile56Thr variant of human lysozyme under a particular set of solution conditions. The quantity of amyloid present was monitored through the change in the fluorescence of thioflavin-T associated with its binding to the fibrils. Values are means of four measurements. Filled squares, protein incubated without seeding; open circles, seeding with 2% by volume of a similar solution in which fibrils of Ile56Thr lysozyme had previously been allowed to form. (Adapted from Morozova-Roche *et al*. 2000.)

variants, fibrils form most readily at low pH or at slightly elevated temperatures (Booth *et al*. 1997; Morozova-Roche *et al*. 2000). Experiments to examine the nature of the amyloidogenic variants (Ile56Thr and Asp67His) show that the structures of the proteins in their soluble native states are similar to that of the wild-type protein and have no obvious perturbations that could explain their tendency to aggregate (Booth *et al*. 1997). But experiments reveal that the two variants are destabilized relative to the wild-type protein to similar extents, although the origin of this instability is different (Canet *et al*. 1999). Thus the Ile56Thr variant is destabilized largely because its folding rate is reduced, whilst the Asp67His is destabilized largely because it unfolds more rapidly. It therefore appears that the decreased protein stability rather than the altered folding kinetics is a common feature of these two variants. In further experiments it was demonstrated that the lower stability of the native state results in the population of a partially folded state that is very similar to the major (α domain) intermediate populated on the folding pathway of the wild-type protein (Canet *et al*. 1999). This finding can be rationalized because the mutations destabilizing the native fold are located in the β domain of the protein, the region that is not highly structured in the predominate intermediate (figure 4).

This observation suggests a mechanism for the formation of amyloid fibrils from the variant lysozymes, in which the partially folded intermediates aggregate in a first step in the formation of the ordered structure found in the fibrils (Booth *et al*. 1997; Canet *et al*. 1999; Morozova-Roche *et al*. 2000). The unfolded region of the intermediate is primarily in the β domain of the protein, suggesting that the aggregation process might be initiated by the intermolecular association of β-strands that are involved in intramolecular interactions in the native structure (figure 5). This proposition is supported by the observation of the ready formation of amyloid fibrils from

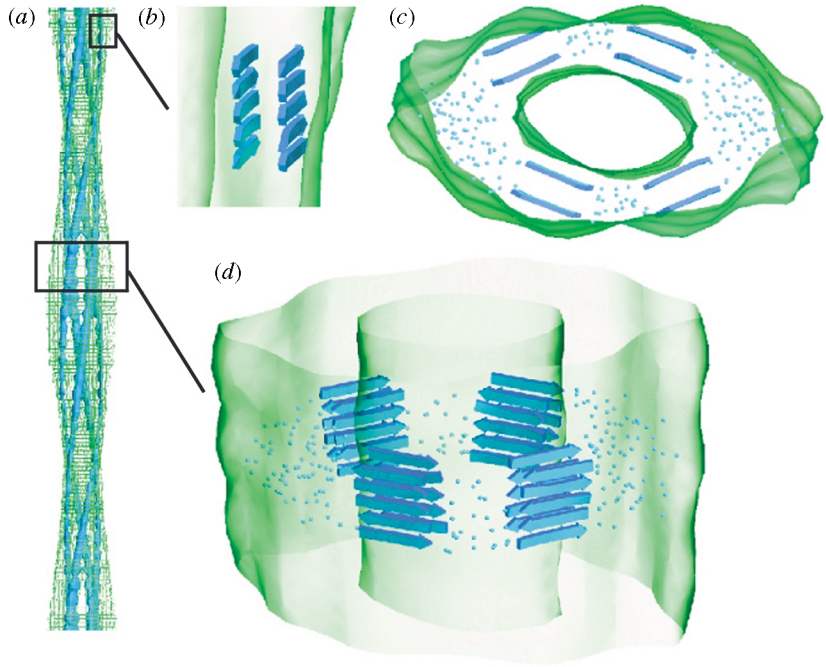

Figure 7. Model of the polypeptide fold in SH3 fibrils. (*a*) Overview of the fibril structure, showing the outer surface as a green mesh and the protofilaments as solid blue surfaces. The ribbon-like protofilaments form the skeleton of the fibril structure. A model for the molecular packing is shown in (*b–d*), with the electron micrograph map as a transparent rendered surface. (*b*) Side view of a single protofilament. (*c*) Cross-section of the fibril and (*d*) slightly tilted side view of the fibril. β-sheets derived from the SH3 structure have been fitted into the map, after opening the β-sandwich fold and reorientating and straightening the strands. The remaining regions of polypeptide sequence are shown as disconnected dots, to indicate the number of residues present but not the conformation. The β-sheets contain a mixture of parallel and antiparallel strands. This particular arrangement is arbitrary and was chosen because it required the least rearrangement of the native β-sheet structure, although there is no evidence that the regions of sheet that are present in the native fold are specifically or uniquely present in the fibrils. The β-sheets fit well into the protofilament density, and the loops provide the right amount of mass to generate the rest of the density. (From Jiménez *et al*. 1999.)

a peptide corresponding to part of the β domain of the homologous hen lysozyme (Krebs *et al*. 2000). In terms of the energy landscape model of protein folding, discussed in §1 and illustrated in figure 2, the mutations destabilize the native state such that, under conditions where the wild-type protein is stable in its native state, the variants may not be. This destabilization can allow the variants to access partially folded states through the fluctuations in structure, inherent in all proteins, under conditions where the wild-type protein cannot. Calculations based on hydrogen exchange protection suggest that the population of partially folded proteins under physiological conditions is nearly 1000 times greater in the variants than in wild-type lysozyme (Canet *et al*. 1999). This conclusion allows one to speculate that the amyloidogenic variants have sufficient stability to fold efficiently so as to escape the quality control mechanisms in the endoplasmic reticulum and to be secreted into extracellular space (Dobson 1999*a*). However, unlike the wild-type protein, they have insufficient stability to remain in their native states under all conditions to which they are exposed. Moreover, it has been speculated that endosomal compartments where the pH is reduced might be important in the formation of amyloid. Under low pH conditions *in vitro* conversion to amyloid fibrils has been found to be particularly facile (Morozova-Roche *et al*. 2000). In addition, *in vitro* experiments have shown that fibril formation is accelerated substantially when solutions are seeded with preformed fibrils (figure 6). Such a mechanism has been suggested as being responsible for the rapid onset of some amyloidoses, and indeed of the infectivity of the prion diseases (Lansbury 1999).

The picture emerging from the studies of lysozyme has many features in common with the results of studies of other amyloidogenic proteins, notably those of transthyretin (Kelly 1998). Of particular importance is the finding that the native state needs to be disrupted to allow fibril formation to occur, and that at least partially unfolded species are accessible under conditions when such fibril formation is rapid. This conclusion is consistent with the observation that most amyloidogenic mutations in the various proteins associated with disease are destabilizing, and that several diseases are associated with fibrils formed by fragments of proteins that are not able to fold into a native-like structure (Kelly 1998). There are, however, many questions about amyloid formation that remain to be answered. For example, is there any common feature in the group of 20 or so proteins that explains why they form fibrils *in vivo* whilst other proteins do not appear to do so? And what is the structure of the fibrils that results in them appearing so similar, regardless of the nature of the protein from which they form? A chance observation during studies of the folding of a small SH3 domain was the stimulant to further work from our laboratory that begins to address both these questions.

3. THE GENERIC NATURE OF AMYLOID STRUCTURE

In studies of the conformation of the SH3 domain from bovine PI3 kinase at low pH, when the protein is in a largely unfolded state, it was found that the protein readily formed a viscous gel. Examination of the gel using electron microscopy revealed the presence of large numbers of fibrils that closely resemble those formed from the proteins associated with amyloid diseases (Guijarro *et al.* 1998). Moreover, the aggregates showed all the other characteristics of amyloid fibrils, and were to all intents and purposes identical to these other structures. This observation prompted us to explore the possibility that similar fibrils could be formed from other proteins by placing them under mildly denaturing conditions that do not immediately result in visible precipitation, and examining the solutions over often prolonged periods of time (Chiti *et al.* 1999*b*). For a range of representative proteins with no known connection with any disease we have been able to find conditions under which conversion occurs into fibrils very similar to those associated with amyloid disease (table 3). We shall refer to these types of structures as 'amyloid fibrils' in future, regardless of whether or not they are associated with disease. The proteins studied included wild-type human lysozyme, which forms fibrils under similar but more destabilizing conditions than the amyloidogenic intermediates (Morozova-Roche *et al.* 2000), and the archetypal globular protein, myoglobin (Fändrich *et al.* 2001). For the latter it is particularly evident that the protein has undergone a substantial conversion from its soluble α-helical form to the aggregated β-sheet conformation found in the fibrils. Such findings, along with a number of related observations, prompted us to conclude that the ability to form amyloid fibrils is not a characteristic associated wholly or primarily with those proteins found to be associated with amyloidoses, but a property that could be common to many or indeed all proteins under appropriate conditions (Guijarro *et al.* 1998; Chiti *et al.* 1999*b*; Dobson 1999*a*).

Table 3. *Conversion of representative globular proteins into amyloid fibrils* in vitro

protein	native structure type	reference
PI3–SH3 domain	β	Guijarro *et al.* (1998)
Fn III domain	β	Litvinovich *et al.* (1998)
acylphosphatase	α/β	Chiti *et al.* (1999*b*)
ADA2h	α/β	Villegas *et al.* (2000)
α-lactalbumin	α + β	C. Redfield *et al.*, unpublished data
lysozyme	α + β	Krebs *et al.* (2000); Morozova -Roche *et al.* (2000)
cytochrome c_{552}	α	Pertinhez *et al.* (2001)
apo-myoglobin	α	Fändrich *et al.* (2001)

The ability to generate samples from a wide variety of protein sequences provides us with an opportunity to study a much wider range of properties of amyloid fibrils than was previously possible. Indeed, the SH3 fibrils turn out to be particularly well ordered when grown slowly under controlled conditions, and in some cases have an extremely regular helical twist. These features of the fibrils have enabled high-resolution cryo-electron microscopic studies of their structure to be carried out, and a relatively detailed electron density map to be obtained (Jiménez *et al.* 1999). This study reveals the fibrils to be composed of four 'protofilaments', wound around a hollow core to generate a twisted hollow tube. From the electron density map a tentative model of a possible arrangement of the β-sheet structure could be generated (figure 7). This model, like several earlier proposals, indicates that the structure of the protofilaments is based on hydrogen bonds between the polypeptide main chain (Sunde & Blake 1997; Jiménez *et al.* 1999). As this feature is common to all polypeptides, it explains how the fibrils from different proteins appear so similar, regardless of the length and sequence of the polypeptide involved. In contrast to the situation in native proteins, we suggest that the side chains are not a strong influence on the basic structure of the protofilaments. Nevertheless, the manner in which the protofilaments pack together to form mature fibrils may well depend significantly on those parts of the polypeptide chain that are not involved directly in the close-packed β-strands (Chamberlain *et al.* 2000). Thus, the fibrils from different peptides and proteins are variations on a common theme. One can speculate that the dimensions of the protofilaments, and the lengths of the β-strands within them, may be determined simply by the physicochemical properties of an unfolded polypeptide chain. Presumably the length of the strands in a regular structure will be a balance between the stabilizing contributions of individual hydrogen bonds and the probability of a turn occurring in the polypeptide chain. It is interesting in this regard that the persistence length of a random coil polypeptide chain is approximately eight to ten residues (Smith *et al.* 1996), a range similar to that of the lengths of individual strands in the amyloid structure—and indeed to the typical lengths of strands formed in the native structures of proteins (Branden & Tooze 1999).

The proposal that amyloid fibrils are a generic structure of polypeptide chains has stimulated us to suggest that the conformational properties of all proteins should be considered in terms of the multiple states that are accessible to such structures (Dobson 1999*b*; Dobson & Karplus 1999). This suggestion is illustrated in figure 8 in a schematic manner. This diagram suggests that the various fates awaiting a polypeptide chain once it has been synthesized in the cell will depend on the kinetics and thermodynamics of the various equilibria between different possible states. Thus, the normal folding process may pass through partially folded states on the route to the fully native state, but the aggregation of these species will be minimized by the presence of molecular chaperones. In addition, if the protein is able to fold rapidly, any partially folded species will have a short lifetime, reducing the probability of intermolecular interactions occurring. Moreover, once folded, the native state is generally a highly compact structure that conceals the polypeptide main chain within its interior. Such a state is protected from aggregation except through the interactions of surface side chains (as is the

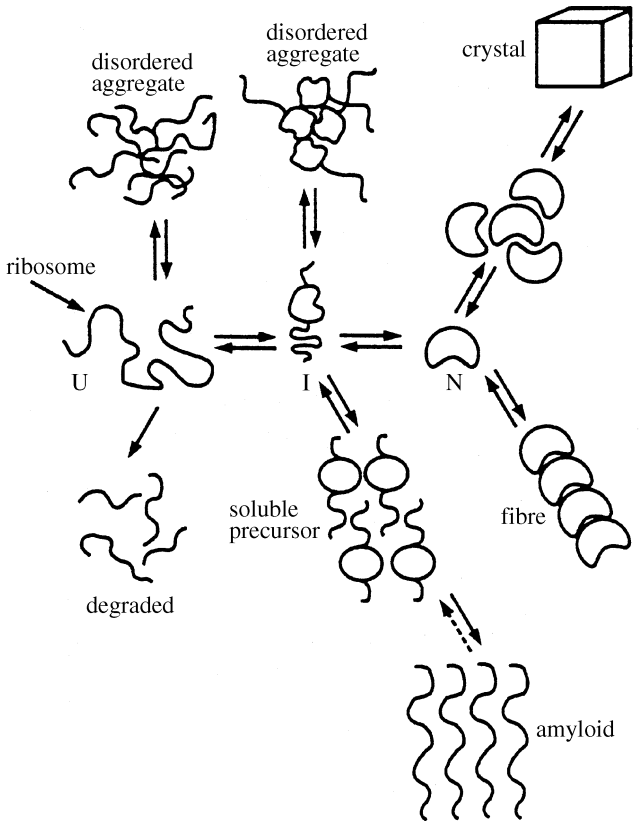

Figure 8. Schematic of some of the states accessible to a polypeptide chain following its biosynthesis. In its monomeric state, the protein is assumed to fold from its highly disordered unfolded state (U) through a partially structured intermediate (I) to a globular native state (N). The native state can form aggregated species, the most ordered of which is a three-dimensional crystal, whilst preserving its overall structure. The unfolded and partially folded states can form aggregated species that are frequently disordered, but highly ordered amyloid fibrils can form through a nucleation and growth mechanism. (From Dobson 1999*b*.)

case, for example, in protein crystals) and is unable to form the strong intermolecular hydrogen bonds associated with the polypeptide backbone. Provided that the native state is maintained under conditions where it remains folded, aggregation to amyloid fibrils will be resisted by the kinetic barrier associated with unfolding, even if the aggregated state is thermodynamically more stable. Importantly, the cooperative nature of protein structures means that virtually none of the polypeptide chain in individual molecules is locally unfolded, and that virtually no molecules in an ensemble are globally unfolded, even though native proteins are only marginally stable relative to denatured ones under normal physiological conditions (Dobson 1999*b*; Dobson & Karplus 1999).

This picture enables us to speculate on the origins of the amyloid diseases from the point of view of the physico-chemical properties of the protein molecules. If the stability or cooperativity of the native state of a protein is reduced, for example by a mutation, the population of non-native states will increase, as discussed above for the amyloidogenic variants of lysozyme. This rise will increase the probability of aggregation, as the concentration of polypeptide chains with at least partial exposure to the external environment will be greater. Whether or not aggregation does occur will depend on the concentration of protein molecules, the intrinsic propensity for a given sequence to aggregate when unfolded, and on the rate of the aggregation process. The fact that formation of ordered amyloid fibrils can be seeded, like the well-studied processes of crystallization and gelation, means that once the aggregation process is initiated it often proceeds very much more rapidly (Harper & Lansbury 1997; Morozova-Roche *et al.* 2000; Krebs *et al.* 2000). In the absence of seeding there can be long 'lag' phases before aggregation occurs (figure 5). This lag can be thought of as arising because the growth of a fibril cannot occur until a 'nucleus' of a small number of aggregated molecules is formed. Such a nucleus can be formed by the local fluctuations in concentration that occur in solution as a result of random molecular motion. When such fluctuations result in a local concentration of molecules above a critical value, the molecules associate with one other to form a species that is sufficiently large to have intrinsic stability, and hence to grow in size by interacting with other molecules in the solution. The act of seeding provides such nuclei to the solution and hence reduces or abolishes the lag phase.

4. A GENERAL MECHANISM OF AMYLOID FORMATION

On this view of the aggregation process, the critical step is the unfolding of the native structure. In the case of most proteins, except the smallest ones discussed at the beginning of this article, unfolding under physiological conditions will not generate the type of highly unfolded states seen in high concentrations of denaturant. Instead the denatured protein will be more stable in a partially collapsed 'intermediate' state that may well resemble the folding intermediates observed in the normal folding process (Dobson *et al.* 1998). The generic nature of the structure and mechanism of amyloid formation suggests that the nature of the residual structure in such intermediates has little direct importance in dictating the structure of the resulting aggregates, although it may indicate which regions of the protein are most likely to be incorporated in the β-sheet segments of the fibrils. For example, in the case of lysozyme it is likely that the β structure in the fibrils is mainly formed from the β domain that is highly unfolded in the intermediate populated under conditions where amyloid forms (Canet *et al.* 1999; Krebs *et al.* 2000). The α domain may be incorporated into the fibrils as a disordered but partially structured helical region connecting the β-strands, as seen in the loop regions of the SH3 structure (Jiménez *et al.* 1999). The intermediates may be important, however, for another reason in that they are generally much more soluble than highly unfolded polypeptides, enabling the concentrations of non-native species to reach those required for nucleation as described above.

This general view of amyloid formation can readily be extended to include the existence of sporadic as well as familial diseases (i.e. those involving mutations that destabilize the protein or provide the unfolded or partially unfolded polypeptide chain with an increased tendency to aggregate) and infectious diseases (for example, if they are

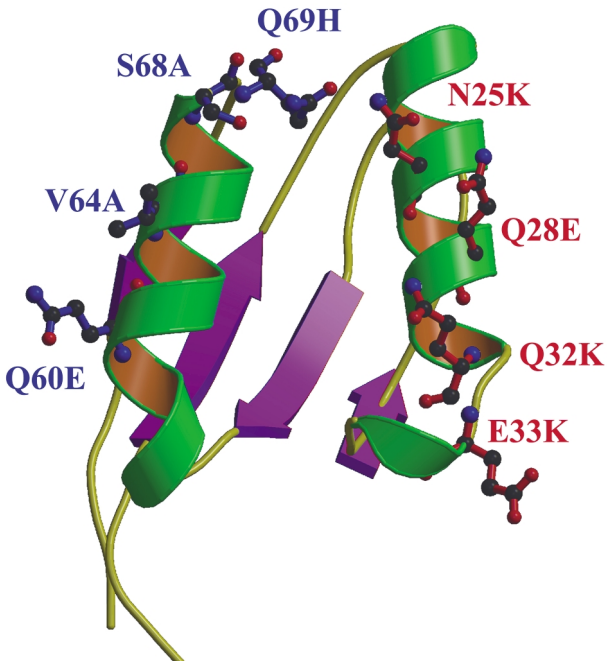

Figure 9. Tertiary structure of the protein ADA2h. Surface mutations in α-helix 1 are shown in red and those in α-helix 2 in blue. Both sets of mutations reduce substantially the propensity of the denatured protein to aggregate, although they have relatively little effect on the structure or stability of the native protein. (From Villegas *et al*. 2000.)

nucleated by seeding with already aggregated species). Such diseases could arise from the loss of the normal control and regulation processes that enable proteins to be maintained in their required states under all conditions in the organism (Chiti *et al.* 1999*b*; Dobson 1999*a*). It is perhaps particularly likely that such control is lost in ageing, and the majority of the cases of sporadic diseases such as Alzheimer's or type II diabetes are associated with old age. It is significant that in a high proportion of elderly people even wild-type transthyretin, which in its mutant forms is associated with familial amyloidosis, is found as amyloid structures in organs such as the heart (Kelly 1998). The exact manner in which this happens is unclear, but it could be as the result of statistical factors (comparable with those observed in lag phases *in vitro*), to the effects of changes in the cellular environment, or to the failure of the normal degradation mechanisms for proteins. Interestingly, many of the diseases in this category involve the deposition of peptide fragments, and the process of degradation in compartments such as lysozomes involves conditions such as low pH that serve to unfold proteins prior to the action of proteases. Such mildly denaturing conditions are particularly favourable for the nucleation and growth of amyloid structures.

The idea that amyloid is a generic form of protein structure leads to the question of why it is only associated with disease (although in yeast and some other fungi it has been suggested that amyloid may play a role in normal behaviour—see Lindquist 1997; Wickner *et al.* 1999). One possible explanation is as follows. The formation of amyloid fibrils is difficult to control, and once formed it is often extremely difficult or impossible to degrade. It is not therefore a favourable material for an efficient living organism to use. It seems, therefore, that biological evolution has managed to select some sequences that are able to fold to compact, globular and soluble forms that resist aggregation and conversion to fibrillar structures, at least when protected in a highly stable and controlled environment (Dobson 1999*a*). These remarkable structures are the native states of proteins that are involved in every process occurring in the cell. The appearance of amyloid in living systems may therefore be associated with mutations that destabilize proteins sufficiently for them to convert into fibrils when the wild-type protein would not, but are sufficiently stable to evade the quality control mechanisms in the cell and to function sufficiently normally to allow the organism to develop and reproduce. Amyloid also appears in old age where evolutionary pressure is reduced after the reproductive life span, and in other conditions such as kuru or bovine spongiform encephalopathy (BSE), which are connected with abnormal practices such as ingestion of tissue from other members of the same species (Prusiner 1997). This conceptual picture enables all of these diseases to be rationalized at least in general terms on a similar basis.

The inherent propensity for proteins—and indeed other molecular species—to aggregate is therefore one of the primary issues that living systems have to control in order to survive. The development of molecular chaperones is one example of a strategy to minimize the effects of aggregation and these molecules are now recognized to be crucial in enabling the cell to produce proteins efficiently and localize them appropriately in the cell (Ellis & Hartl 1999). The evolution of sequences with the very special ability to form cooperative globular structures, and of the control mechanisms to maintain these molecules under appropriate conditions, is undoubtedly an even more fundamental mechanism of avoiding aggregation (Dobson 1999*a*). Just as efficient computers need to be small to enable electrons to move rapidly between components, so living cells need to be small and densely packed to allow molecular diffusion to transfer information efficiently. It is truly a remarkable achievement of evolution to generate in cells solutions of macromolecules at concentrations in excess of $300\,\mathrm{mg\,ml^{-1}}$, comparable with those of some crystalline solids, without non-specific interactions or aggregation (Luby-Phelps 1994; Minton 2000). It is interesting to speculate, however, that the material rejected by biology could be useful in modern technology. The structure of amyloid, with organization on a nanometre scale, suggests that it might have important applications as an almost infinitely functionalizable nanomaterial. Indeed, it has already been shown that non-biological groups with novel optical properties can be introduced into peptide-based fibrils, demonstrating the viability of this proposal (MacPhee & Dobson 2000).

The ability of natural proteins to form amyloid structures does not violate the crucial hypothesis that a protein sequence codes for a single fold (Anfinsen 1973). The nature of amyloid is that it is not coded for by the sequence, as it is formed as a consequence of interactions involving the common polypeptide backbone of all proteins. Its rate and ease of formation will of course depend on the sequence, both as a consequence of the

readiness for different side-chains to pack together within the structure, and as a consequence of the solubility and stability of the sequence in solution (Chiti *et al.* 2000*b*). In this sense the formation of amyloid fibrils can be likened to crystallization. Virtually all molecules crystallize (including, for example, all the individual amino acids), although they differ in their readiness so to do. In general, biology has not used crystalline proteins, except in rare instances such as insulin stored in the pancreas (Dodson & Steiner 1998). It is the side-chains, however, that code for the specific fold of globular proteins by their ability to pack together in a unique manner to form compact globular structures. Furthermore, it is not necessary for an alternative protein fold to be present to allow amyloid to be formed. We believe that the structures of amyloidogenic intermediates do not define the structure of the resulting amyloid fibrils, as discussed above, but that within the ensemble of conformations that comprise unfolded or partially folded states, structures are accessible that can aggregate through the formation of β-sheets. There is therefore no need for the sequence to code for a specific amyloidogenic intermediate in order for fibrils to develop.

5. LOOKING TO THE FUTURE

The generic picture of amyloid structure and the mechanism of its formation provides a conceptual framework for linking together the various pathological conditions associated with deposition of this material, and hence to suggest possible general approaches to the prevention or treatment of the whole family of amyloid diseases. From the viewpoint of physical chemistry, the origin of enhanced amyloidogenicity is either a reduced stability of the native state, or the increased tendency to aggregate of any accessible populations of unfolded or partially folded species. Both of these factors can, as we have discussed, result from single amino-acid mutations. One therapeutic strategy is, therefore, to increase the stability of the protein involved. This approach has been explored in detail for transthyretin, where substrate analogues that stabilize the native state have been shown to reduce the tendency for amyloidogenic mutants of the protein to form fibrils (Klabunde *et al.* 2000). Studies of the small protein acylphosphatase have shown that this is likely to be a general method of reducing the tendency to form amyloid fibrils under conditions where the native state of a protein has low stability (Chiti *et al.* 2000*a*). Investigation of acylphosphatase and ADA2h (Villegas *et al.* 2000; Chiti *et al.* 2000*b*) has also indicated that mutations that reduce the aggregation tendency of the denatured state are very effective at inhibiting fibril formation (figure 9). Indeed, with acylphosphatase, rates of aggregation can be reduced by a factor of 1000 by single changes of amino-acid residues at positions in the protein that do not perturb significantly the stability or functional behaviour of the protein (Chiti *et al.* 2000*b*). These findings suggest that a strategy such as gene therapy could in principle be viable to inhibit an amyloid disease by specific modification of the sequence of the protein whose aggregation is the primary origin of the clinical symptoms. Importantly, it has been demonstrated that liver transplants, which remove the origin of variant proteins associated with amyloidogenesis, can result in remission of disease and gradual disappearance of the fibrils from the body (Stangou *et al.* 1999).

In addition to these strategies, therapeutic methods have been developed that are designed either to inhibit amyloid formation by blocking fibril growth (Soto *et al.* 1998) or to inhibit serum amyloid protein (SAP), a protein that binds to fibrils *in vivo* and protects them from succumbing to the normal degradative processes (Gillmore *et al.* 1997). Such approaches are very promising, and the enhancement of our knowledge of the mechanism of fibril growth and degradation should help enormously to design effective therapeutic strategies. The fundamental knowledge that can come from the ability to probe the aggregation properties of a range of sequences, rather than just those identified in association with recognized diseases, should enable the role of particular interactions and specific structural motifs in these processes to be explored. This knowledge could result in design strategies that are significantly more general and potentially more effective. Given that many of the amyloid diseases are associated with old age, the need for novel approaches to drug design will be increasingly important in the future. One can speculate that as the human life span increases, the number of cases of the known diseases is likely to increase substantially, and indeed that novel diseases associated with the aggregation of proteins not so far linked to clinical symptoms might become significant in ageing populations.

As well as the significance of this work in understanding the relationship between folding, misfolding and disease, there are other consequences of the recent findings about amyloid formation that have been discussed in this paper. The ability to understand the aggregation behaviour of proteins in more detail, and to develop ways of controlling it, has tremendous possible significance for increasing the use of proteins for commercial purposes, for example as catalysts of specific reactions. In addition, the ability to control aggregation phenomena has considerable importance in terms of fundamental studies of their properties. In the context of structural biology, for example, control of solubility is of critical importance in inducing efficient crystallization of proteins for X-ray studies, and in generating increased solubility for NMR studies. In addition, the nature of amyloid fibrils suggests that they could form the basis of almost infinitely functional nanomaterials that are biocompatible and highly stable over a wide range of conditions (MacPhee & Dobson 2000). The development of these various lines of enquiry from observations that were initially somewhat fortuitous (Guijarro *et al.* 1998) illustrates the importance of interdisciplinary research at the interfaces of the physical, biological and medical sciences.

The work described in this article has involved a wide range of students, post-doctoral workers and colleagues in the University of Oxford and elsewhere. The names of many of these people, without whom none of this work would have been possible, are largely to be found in the papers cited in this article. I am particularly indebted to John Ellis for critical reading of this manuscript and for assisting in the preparation of table 1. The Oxford Centre for Molecular Sciences is supported by the Biotechnology and Biological Sciences Research Council, the Engineering and Physical Sciences Research Council and

the Medical Research Council. The research of C.M.D. is also supported by the Wellcome Trust.

REFERENCES

Anfinsen, C. B. 1973 Principles that govern the folding of protein chains. *Science* **181**, 223–230.

Baker, D. 2000 A surprising simplicity to protein folding. *Nature* **405**, 39–42.

Booth, D. R. (and 10 others) 1997 Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**, 787–793.

Branden, C. & Tooze, J. 1999 *Introduction to protein structure*, 2nd edn. New York: Garland Publishing.

Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. 1995 Funnels, pathways and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195.

Callender, R. H., Dyer, R. B., Gilmashin, R. & Woodruff, W. H. 1998 Fast events in protein folding. *A. Rev. Phys. Chem.* **49**, 173–202.

Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V. & Dobson, C. M. 1999 Mechanistic studies of the folding of human lysozyme and the origin of amyloidogenic behavior in its disease related variants. *Biochemistry* **38**, 6419–6427.

Chamberlain, A. K., MacPhee, C. E., Zurdo, J., Morozova-Roche, L. A., Hill, H. A. O., Dobson, C. M. & Davis, J. J. 2000 The ultrastructural organisation of amyloid fibrils by atomic force microscopy. *Biophys. J.* **79**, 3282–3293.

Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M. & Dobson, C. M. 1999*a* Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**, 1005–1009.

Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. 1999*b* Designing conditions for *in vitro* formation of amyloid protofilaments and fibrils. *Proc. Natl Acad. Sci. USA* **96**, 3590–3594.

Chiti, F., Taddei, N., Bucciantini, M., White, P., Ramponi, G. & Dobson, C. M. 2000*a* Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J.* **19**, 1441–1449.

Chiti, F., Capanni, C., Taddei, N., Stefani, M., Ramponi, G. & Dobson, C. M. 2000*b* Specific regions of a protein determine the kinetics of aggregation and amyloid formation. (In preparation.)

Dill, K. A. & Chan, H. S. 1997 From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10–19.

Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. 2000 Understanding protein folding via free energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**, 331–339.

Dobson, C. M. 1999*a* Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**, 329–332.

Dobson, C. M. 1999*b* How do we explore the energy landscape for folding? In *Simplicity and complexity in proteins and nucleic acids* (ed. H. Fraunfelder, J. Deisenhofer & P. G. Wolynes), pp. 15–37. Berlin: Dahlem University Press.

Dobson, C. M. & Fersht, A. R. (eds) 1995 Protein folding. *Phil. Trans. R. Soc. Lond.* B **348**, 1–119.

Dobson, C. M. & Hore, P. J. 1998 Kinetic studies of protein folding using NMR spectroscopy. *Nature Struct. Biol.* **5**, 504–507.

Dobson, C. M. & Karplus, M. 1999 The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**, 92–101.

Dobson, C. M., Evans, P. A. & Radford, S. E. 1994 Understanding protein folding—the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31–37.

Dobson, C. M., Sali, A. & Karplus, M. 1998 Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37**, 868–893.

Dodson, G. & Steiner, D. 1998 The role of assembly in insulin's biosynthesis. *Curr. Opin. Struct. Biol.* **8**, 189–194.

Dyson, H. J. & Wright, P. E. 1998 Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Struct. Biol.* **5**, 499–503.

Ellis, R. J. & Hartl, F. U. 1999 Principles of protein folding in the cellular environment. *Curr. Opin. Struct. Biol.* **9**, 102–110.

Evans, P. A. & Radford, S. E. 1994 Probing the structure of folding intermediates. *Curr. Opin. Struct. Biol.* **4**, 100.

Fändrich, M., Fletcher, M. A. & Dobson, C. M. 2001 Formation of amyloid fibrils from myoglobin. (Submitted.)

Fersht, A. R. 1997 Nucleation mechanisms in protein folding, *Curr. Opin. Struct. Biol.* **7**, 3–9.

Fersht, A. R. 1999 *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*. New York: W. H. Freeman.

Gething, M.-J. & Sambrook, J. 1992 Protein folding in the cell. *Nature* **355**, 33–45.

Gillmore, J. D., Hawkins P. N. & Pepys, M. B. 1997 *Brit. J. Haematol.* **99**, 245–256.

Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. 1998 Amyloid fibril formation by an SH3 domain. *Proc. Natl Acad. Sci. USA* **95**, 4224–4228.

Harper, J. D. & Lansbury, P. T. 1997 Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *A. Rev. Biochem.* **66**, 385–407.

Jiménez, J. L., Guijarro, J. I., Orlova, E., Zurdo, J., Dobson, C. M., Sunde, M. & Saibil, H. R. 1999 Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* **18**, 815–821.

Karplus, M. 1997 The Levinthal paradox: yesterday and today. *Fold Des.* **2**, S69–S75.

Kelly, J. W. 1998 The alternative conformations of amyloid proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8**, 101–106.

Klabunde, T., Petrassi, H. M., Oza, V. B., Raman, P., Kelly, J. W. & Sacahetini, J. C. 2000 Rational design of potent human transthyretin amyloid disease inhibitors. *Nature Struct. Biol.* **7**, 312–321.

Krebs, M. R. H., Wilkins, D. K., Chung, E. W., Pitkeathly, M. C., Chamberlain, A., Zurdo, J., Robinson, C. V. & Dobson, C. M. 2000 Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the β-domain. *J. Mol. Biol.* **300**, 541–549.

Lansbury, P. T. 1999 Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl Acad. Sci. USA* **96**, 3342–3344.

Linquist, S. 1997 Mad cows meet *psi*-chotic yeast: the expansion of the prion hypothesis. *Cell* **89**, 495–498.

Litvinovich, S. V., Brew, S. A., Aota, S., Haudenschild, S. K. C. & Ingham, K. C. 1998 Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *J. Mol. Biol.* **280**, 245–258.

Luby-Phelps, K. 1994 Physical properties of cytoplasm. *Curr. Biol.* **6**, 3–9.

MacPhee, C. E. & Dobson, C. M. 2000*a* Chemical dissection and reassembly of amyloid fibrils formed by a peptide fragment of transthyretin. *J. Mol. Biol.* **297**, 1205–1217.

MacPhee, C. E. & Dobson, C. M. 2000*b* Formation of mixed fibrils reveals the generic nature and potential application of protein amyloid structures. *J. Am. Chem. Soc.* **122**, 12 707–12 713.

Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. 1989 Mapping the transition-state and pathway of protein folding by protein engineering. *Nature* **340**, 122–126.

Minton, A. P. 2000 Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **10**, 34–39.
Morozova-Roche, L. A., Jones, J. A., Noppe, W. & Dobson, C. M. 1999 Independent nucleation and heterogeneous assembly of structure during folding of equine lysozyme. *J. Mol. Biol.* **289**, 1055–1073.
Morozova-Roche, L. A., Zurdo, J., Spencer, A., Noppe, W., Receveur, V., Archer, D. B., Joniau, M. & Dobson, C. M. 2000 Amyloid fibril formation and seeding by wild type human lysozyme and its disease related mutational variants. *J. Struct. Biol.* **130**, 339–351.
Pertinhez, T. A., Bouchard, M., Tomlinson, E. J., Wain, R., Smith, L. J., Ferguson, S. J. & Dobson, C. M. 2001 Amyloid fibril formation by a helical cytochrome. (In preparation.)
Perutz, M. F. 1999 Glutamine repeats and neurodegenerative disease: molecular aspects. *Trends Biochem. Sci.* **24**, 58–63.
Plaxco, K. W. & Dobson, C. M. 1996 Time-resolved biophysical methods in the study of protein folding. *Curr. Opin. Struct. Biol.* **6**, 630–636.
Plaxco, K. W., Simons, K. T. & Baker, D. 1998 Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.
Prusiner, S. 1997 Prion diseases and the BSE crisis. *Science* **278**, 245–251.
Radford, S. E. & Dobson, C. M. 1999 From computer simulations to human disease: emerging themes in protein folding. *Cell* **97**, 291–298.
Shortle, D. 1996 The denatured state and its role in protein stability. *FASEB J.* **10**, 27–34.
Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. 1996 The concept of a random coil. Residual structure in peptides and denatured proteins. *Fold. Des.* **1**, 95–106.
Soto, C., Sigurdsson, E. M., Morelli, L., Kumar, R. A., Castano, E. M. & Frangione, B. 1998 β-Sheet breaker peptides inhibit fibrillogenesis in a rat brain model of amyloidosis: implications for Alzheimer's therapy. *Nature Med.* **4**, 822–826.
Stangou, A. J., Hawkins, P. N., Booth, D. R., O'Grady, J., Jewitt, D., Rela, M., Pepys, M. B. & Heaton, N. D. 1999 Liver transplantation for non-Met30 TTR associated familial amyloid polyneuropathy. *Hepatology* **30**, 576.
Sunde, M. & Blake, C. C. F. 1997 The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50**, 123–159.
Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B. & Blake, C. C. F. 1997 Common core structure of amyloid fibrils by synchrotron X-ray diffractions. *J. Mol. Biol.* **273**, 729–739.
Tan, S. Y. & Pepys, M. B. 1994 Amyloidosis. *Histopathology* **25**, 403–414.
Thirumalai, D. & Klimov, D. K. 1999 Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. *Curr. Opin. Struct. Biol.* **9**, 197–207.
Thomas, P. J., Qu, B. H. & Pederson, P. L. 1995 Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20**, 456–459.
Villegas, V., Zurdo, J., Filimonov, V. V., Aviles, F. X., Dobson, C. M. & Serrano, L. 2000 Protein engineering as a strategy to avoid formation of amyloid fibrils. *Protein Sci.* **9**, 1700–1708.
Wickner, R. B., Edskes, H. K., Maddelein, M. L., Taylor, K. L. & Moriyama, H. 1999 Prions of yeast and fungi—proteins as genetic material. *J. Biol. Chem.* **274**, 555–558.

Discussion

B. E. P. Swoboda (*Department of Biological Sciences, University of Warwick, UK*). Do you think that all newly synthesized proteins are protected by chaperones or are there some 'Anfinsen' proteins that do not require them?

C. M. Dobson. I think it is likely that molecular chaperones of one type or another are involved with the folding of at least the majority of proteins. It is clear, however, that in the case of some chaperones such as GroEL, only a relatively small proportion of proteins use them to assist their folding (Ellis 2000).

M. Joniau (*University of Leuven, Belgium*). You showed us a multicoloured picture of the β-stack structure of a fibril with a left-handed turn. Then you showed a picture of the insulin fibril, which has a right-handed turn. How can one reconcile these two structures?

C. M. Dobson. Many of the amyloid fibrils observed in electron micrograph studies appear to be made up of an assembly of protofilaments that are twisted together. The handedness of such structures cannot be revealed directly from such images, but more detailed studies can define this property. It appears that most of the fibrils studied so far are left-handed (Chamberlain *et al.* 2000), but this was not known when some of the models were constructed (Jiménez *et al.* 1999).

Anonymous. I have two questions about amyloid fibrils. What is their concentration *in vitro*? Do you observe the formation of mixed fibrils?

C. M. Dobson. Fibrils form under a variety of conditions, depending on the protein. As an example, the fibrils shown in figure 3 were grown from a solution of $10\,\mathrm{mg\,ml^{-1}}$ of the Ile56Thr variant of human lysozyme at pH 2.0 and 37 °C. The fibrils can, however, be concentrated by centrifugation or other means to much higher levels. We have shown that mixed fibrils can be formed under some circumstances (MacPhee & Dobson 2001), but we do not yet have enough evidence to comment on the generality of this finding. We are presently investigating this phenomenon in detail.

C. Weissmann (*Medical Research Council Prion Unit, Imperial College School of Medicine at St Mary's Hospital, London, UK*). Are there effects of crowding on water activity *in vivo*?

C. M. Dobson. The effects of crowding on the thermodynamic properties of solutions are dramatic. For example, in accordance with crowding theory, the association constants of interacting macromolecules are increased 100- to 1000-fold by the addition of crowding agents that mimic the total concentrations of macromolecules inside cells. This effect has long been known, but has been largely neglected by biochemists and molecular biologists—it deserves much more study (see Minton 1998).

Additional references

Ellis, R. J. 2000 Chaperone substrates inside the cell. *Trends Biochem. Sci.* **25**, 210–212.
Minton, A. P. 1998 Molecular crowding: analysis of effects of high concentrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion. *Methods Enzymol.* **295**, 127–149.